

Using Bilingual Test to Assess Malaysian Limited English Proficient Students' Mathematics Achievement

(Ujian Dwibahasa untuk Menilai Pencapaian Matematik Pelajar yang Terhad Penggunaan Bahasa Inggeris)

S.KANAGESWARI D/O SUPPIAH SHANMUGAM* & ONG SAW LAN

ABSTRACT

Language being a measure of test construct inhibits limited English proficiency (LEP) students' performance. This study examined the effects of using bilingual (English and Malay) and English-only test booklets among Malaysian LEP students. Their mathematics achievement was investigated by examining the equivalence of the two test booklets by linking the scores using RAGE-RGEQUATE Version 3.22. The 2,021 students came from 34 schools from the states of Penang and Perak. 1049 sat for the bilingual test while 972 for the English-only test which were spirally administered using random group design. Both booklets contain the same 40 multiple-choice Mathematics items obtained from timss 1999 and 2003 released items. Results indicated that a score difference of one unit between the two tests was due to equating error and therefore the bilingual test was not easier for the LEP students, but helpful as they still used the adapted language version to ease understanding. However, they still preferred to answer in English as they were more familiar with the mathematical terminologies that were used. A better alternative is to simplify the linguistic complexity so that language of instruction matched language of assessment.

Keywords: Bilingual testing, Limited English proficiency, Test linking, Validity

ABSTRAK

Bahasa merupakan satu konstruk ujian yang membantut pencapaian pelajar kurang mahir (PKM) dalam Bahasa Inggeris (PKM). Kajian ini menyiasat kesan penggunaan ujian dwibahasa (Bahasa Inggeris dan Bahasa Melayu) dan ujian Matematik dalam Bahasa Inggeris dalam kalangan PKM di Malaysia. Pencapaian matematik ditinjau dengan menyelidik kesetaraan kedua-dua ujian dengan mengaitkan skor dengan menggunakan RAGE-RGEQUATE Versi 3.22. Seramai 2,021 pelajar dari 34 buah sekolah dari negeri Pulau Pinang dan Perak dikaji. Seramai 1,049 pelajar menduduki ujian dwibahasa manakala 972 menduduki ujian dalam Bahasa Inggeris yang ditadbir secara berlingkar dengan menggunakan reka bentuk kumpulan rawak. Kedua-dua ujian mengandungi 40 item Matematik aneka pilihan yang diperolehi daripada TIMSS 1999 dan TIMSS 2003. Dapatan menunjukkan bahawa perbezaan skor satu unit di antara kedua-dua ujian adalah disebabkan oleh ralat penyamaan dan justeru itu, ujian dwibahasa bukan lebih senang bagi PKM. Namum, pengadaptasian ke Bahasa Melayu tetap membantu mereka untuk memahami soalan. Dalam pada itu, pelajar lebih selesa menjawab dalam Bahasa Inggeris kerana mereka lebih fasih dengan terminologi matematik dalam Bahasa Inggeris yang digunakan semasa mengajar di kelas. Satu alternatif yang lebih sesuai adalah untuk meringkaskan unsur linguistik supaya bahasa pengantar di kelas padan dengan bahasa pengujian.

Kata Kunci: Ujian dwibahasa, Kurang mahir dalam Bahasa Inggeris, Kebocoran ujian, Kesahan

INTRODUCTION

The English-Medium Instruction policy was enacted on 19th July 2002 which mandated the English Language be made the medium of instruction for Mathematics, Science and Technical subjects from 2003. Since Malaysians do not come from English speaking background, teaching students in a language that they are not proficient in threatens test validity. This is because the language barrier introduces construct-irrelevant variance, turning language into an extraneous construct that the test measures and is reflected in the test score (Messick 1995). The emergence of this unintended construct gives rise to unfairness in testing Malaysian limited English proficient (LEP) students.

TEST ACCOMMODATION FOR LEP STUDENTS

For LEP students, the language barrier that introduces construct-irrelevant variance to the test scores needs to be removed. Test accommodation assists them to overcome the language barriers that obstruct their direct access to the test content and allows them to process the content in the target language unlike their peers who have already mastered the language (Center for Excellence and Equity in Education 2005). According to Sireci (1997), test accommodations are not intended to put one group at an advantage but to ensure fairness for all so that each student regardless of their different linguistic backgrounds have equal opportunity to fairly demonstrate their skills. The goal is to address the issue of equity and validity in testing students who may otherwise perform poorly due to certain personal traits like language impediments which may result in inaccurate and invalid interpretation of scores (ETS 2002).

BILINGUAL TESTING

Bilingual test gives provision for the original test items to be translated into the students' dominant language and as such, it removes the unnecessary language barrier (AERA et al. 1999). In the context of the Malaysian education system, the bilingual assessment was administered to cushion the sudden implementation of English as the language

of instruction, (Ainan Abdul Samad 2003; Ministry of Education 2004). This effort was taken as there is the perennial issue of questionable levels of language proficiency among Malaysian students since English is learnt and not acquired (Ain Nadzimah & Chan 2003).

Duncan et al. (2005) explored test accommodation of using dual-language test booklet among LEP students. They used items in English and items that were adapted into the Spanish language. To evaluate the effectiveness of the test booklet, they randomly assigned the bilingual test booklet and the English-only booklet to Spanish students and native English speakers who were the control group. They discovered that the bilingual test booklets were well received by students and was related to their English proficiency. The bilingual test booklets were used a 'comprehension check' to understand the questions since the meanings shifted across the two languages. Even though their findings revealed that 85% of Spanish speaking LEP students found the bilingual booklet useful, they emphasised on the need for further exploration on bilingual test booklet. This perspective was also shared by Wilde (2007) who believed that the mathematics bilingual test appeared to be effective but a better alternative was to simplify the linguistics features of the test items.

Findings from another study conducted by Hofstetter (2003) revealed that the LEP students who received the mathematical instruction in English did not do well as when the items were written in their native language which was Spanish. Similar performance was also recorded among the non-LEP students. However among LEP students who received the mathematical instruction in Spanish, they performed better when the items were translated to Spanish. These observations highlight that the language of instruction should match the language of assessment regardless of students' degree of language proficiency. She deduced that for LEP students, it is imperative for the language of instruction to match the language of assessment. This observation, however, did not apply to the non-LEP students in her study. She also pointed out that simplifying the English language benefitted the LEP students.

In a study conducted by Abedi et al. (2006), all LEP and non-LEP students sat for a student

questionnaire and a reading proficiency test which served as a guide to gauge their language proficiency as high, medium or low. Three different test booklets (standard, bilingual and linguistically modified) were randomly administered to all the students within a class. All the students were matched for their mathematical ability based on their mathematics scores from a mathematics achievement test. Contrary to the earlier results, their findings revealed that LEP students who received the dual-language test booklet did not exhibit differences in their performance when compared to the LEP students who received the standard test, despite indicating high preference for the dual-language test booklet. However, both LEP and non-LEP students displayed better performance when attempting linguistically modified items. Their study highlighted that linguistics simplification benefitted particularly the LEP students more than bilingual test booklet.

MATHEMATICAL LEARNING AMONG MALAYSIAN STUDENTS

Language is a significant factor in mathematical learning. The findings of a study conducted by Lim (2002) in Malaysia indicated that Malay pupils from Chinese medium schools did not do well in mathematics particularly in word problems. The cause was traced back to their poor grasp in their non-mother tongue language which was the Chinese language. Students should master a language well enough to understand the 'language' component embedded in the mathematical concept before solving the problems. The determinant factor was not ethnicity but, the language and the mathematical learning culture of the school as all Chinese and Malay medium schools followed the same Mathematics curriculum.

STATEMENT OF PROBLEM

As language is instrumental in any assessment, all assessments are measures of language skills and as such, language becomes a measure of that test construct (AERA, APA & NCME, 1999). Therefore, the test scores may not accurately demonstrate students' performance as factors like limited

English proficiency influence their performance. In the context of this research, English was the language of instruction for Mathematics and to facilitate the LEP students, Bahasa Malaysia being the national language was used as one of the languages of assessment. However, translating items may confuse students who have learnt these concepts in English as they may not be familiar with the mathematical terminologies in the translated language (Abedi 2004). In addition, translation does not assure equivalent test forms as test content, test constructs, item difficulty and, test reliability and validity can differ (AERA, APA & NCME, 1999).

Therefore, at one hand, translating may pose a potential threat to test validity as the equivalence of the test items in the two different languages may not be similar while on the other hand, translating as a test accommodation allows the students to take the test in their more proficient language. This in return, creates a testing condition where language does not interfere LEP students and yet, provides the same opportunity that the non-LEP students already have.

RESEARCH SIGNIFICANCE

Language should not inhibit any child from gaining access to a fair education, and assessment should be reliable and valid for students of any language background. Since this research is targeted at the LEP students, it is of great significance as the study explores the mathematics achievement of mainstream Malaysian students who are linguistically impoverished due to home background or rural-urban divide. Being LEP in Malaysia is different than being LEP in countries other than Malaysia. In countries other than Malaysia where most of the literature of LEP distinction originates, there is an influx of minority groups of LEP students with diverse ethnic backgrounds residing as immigrants amidst the mainstream non-LEP students (Abedi et al. 2006; Francis et al. 2006; Pena 2007). In Malaysia however LEP students compose the mainstream students who follow the same Malaysian curriculum whose English proficiency is limited due to factors that stretch beyond their nationality.

RESEARCH AIM AND RESEARCH OBJECTIVE

This study is aimed at examining the validity of using bilingual test to assess Form Two Malaysian LEP students' mathematics achievement in order to compare LEP students' mathematics achievement in the English-only and bilingual tests. Therefore, the objective is to determine the equivalence of the English-only test booklet and the bilingual test booklet by linking the test scores using the equipercentile method. By doing so, the LEP students performance in both tests can be compared.

METHODOLOGY

Steps undertaken in this research involved constructing the bilingual and the English-only tests, identifying LEP students, administering the two tests before conducting test linking to determine the LEP students' mathematics achievement.

The research design involved administering the two test booklets to the LEP student sample by using random equivalent group design with spiral administration. LEP students were identified by using teacher judgment which was done by the English teachers who had been teaching the students.

MATHEMATICS TEST

To develop and validate the Mathematics test booklets, three committees of teacher-experts were formed. The first committee consisted of three Form One and Form Two bilingual Malaysian Mathematics teacher whose role was to select items that were defined in the Form One and Form Two Mathematics syllabi. Based on the selected items, a test specification table was built.

The second committee was a committee of three teacher-translators who were proficient in both English and Malay languages, and possessed assessment literacy such as knowledge of test construction. Their role was to translate and adapt the English items into the Malay language. They were given a duration of two weeks to adapt the items. They were encouraged to use simple language and were cautioned not to change the meaning. Some of the suggestions made were

that the currency and the names of people used be changed to suit local context.

The third committee consisted of another independent verification committee of 20 Form One and Form Two bilingual Malaysian Mathematics teachers. Their role was to establish the content validity by responding to statements in a questionnaire provided to them. The questionnaire used a five point Likert rating scale with a response format of agreement (1= 'very strongly disagree' to 5= 'very strongly agree'). The teachers circled the number that best expressed their views on how well the items fit the content domain.

All the teachers in these three committees were from the three main ethnic groups of Malay, Chinese and Indian, with experience of at least 5 years in teaching Form One and Form Two Mathematics. They were experienced bilingual Mathematics teachers who were proficient in both languages, possessed the mathematical content knowledge and assessment literacy, and were familiar with the Malaysian culture. Based on the suggestions from the third committee, 58 items were pilot tested.

MATHEMATICS TEST BOOKLETS DESIGN

The items were pilot tested. Items that displayed a negative point-biserial, discrimination index that was less than 0.2, and item fit that exceeded 1.20 (Thurlow et al. 2000) were rejected. From 58 items, 40 items were selected to construct the English-only and bilingual test booklets.

The English-only test consisted of two parts. The first part was on the students' particulars like race and gender while the second part consisted of 40 multiple-choice questions (MCQ) that formed the Mathematics test. The Malay and English bilingual test booklet consisted of three parts. The first part was on students' particulars. The second part composed of the same 40 MCQ with the inclusion of the Malay language version. For each item and the accompanying diagrams, the Malay language version appeared immediately after the English form, in a square parenthesis using bold italic print of the same font size. The third part addressed the usefulness of the Malay version to solicit feedback about the usefulness of the bilingual version specifically the inclusion of the Malay version.

PARTICIPANTS

Form Two students (14 years old) formed the participants of this study. LEP students were identified by using teacher judgment. The English language teachers classified their students as LEP or non-LEP. Having taught their students for at least ten months and having assessed them at least twice on an individual basis for the school-based English Oral Test, gives them sufficient information to make a credible judgment about their students' English language proficiency.

A total of 2,021 LEP students and 2,747 non-LEP students were identified from 4,768 students from 34 schools in Penang and Perak, with 2,399 students who sat for bilingual test and 2,369 for the English-only test. Altogether 1,049 LEP students sat for the bilingual test while 972 sat for the English-only test and 1,350 non-LEP students sat for the bilingual test while 1,397 sat for the English-only test. Six form two classes were randomly selected in all the schools and in each class, all the students were selected.

TEST ADMINISTRATION AND TEST SCORING

The students were briefed for five minutes based on the written instructions that the teacher read aloud. They are given an hour to answer the questions. Extra ten minutes were given to answer the part C of the bilingual booklets and after which, all the booklets were collected. The items were scored dichotomously. The correct response was given the score '1' while an incorrect response was scored as '0'. Unanswered items were assigned score '0'.

TEST LINKING

The scores of the two tests were linked as the differences in the scores could be due to the difficulties of the two test booklets, apart from the differences in the students' ability. After linking, the difference in the performance that exists between the two groups is the result of the differences in the difficulty of the two tests (Kolen & Brennan 1995). Test linking was executed by using a software, RAGE-RGEQUATE (Version 3.22) by Kolen (2005) which used the postsmoothed equipercentile method for random group. The bilingual test scores were determined by identifying the scores of the

English test that shared the same percentile rank. By doing so, students' mathematics achievement in the English-only booklet were compared to the bilingual booklet.

ITEM PERCENTAGE FOR THE UTILITY OF THE MALAY TRANSLATION

Based on the students' responses for the third section in the bilingual booklet on the utility of the bilingual version, the percentage of students who used the Malay translation for each item was calculated. Students' written comments were also documented.

FINDINGS

At test level, test statistics were used to compare the two tests. The mean values for item difficulty (Bilingual= 0.57, English-only= 0.58), Item Discrimination (Bilingual= 0.57, English-only= 0.56), Point-biserial (Bilingual= 0.50, English-only= 0.49), and SEM (from KR20) (Bilingual= 2.62, English-only= 2.61) differed by one unit. The KR20 Alpha index (0.92) and KR21 (0.91) were the same for both tests. By comparing the mean values of item difficulty, item discrimination, point-biserial and the KR20 index, it can be deduced that overall at the test level, both the bilingual test and English-only test have similar test characteristics. Therefore at test level, the two tests were comparable and also provided evidence of psychometric equivalence.

Table 1 shows the descriptive statistics for the LEP students in the English-only test and the bilingual test. The mean total score for the bilingual test is higher ($M=17.97$) than the mean total score in the English-only test ($M=17.59$) with a difference of 0.38. This indicates that the LEP students exhibited a higher mathematics achievement in the bilingual test than in the English-only test.

TABLE 1. Central moments for LEP students' total score in both tests

	Bilingual test (n=1049)	English-only (n=972)	test
m	17.97	17.59	
sd	8.22	7.96	
Sk	0.41	-0.58	
Ku	-0.68	0.55	

From Figure 1, it can be noted that the mathematics achievement of students in the score range of zero to ten is almost similar in both tests. The middle ability group of students who obtained scores from 11 to 20 seemed to have performed slightly better in the bilingual test than in the English-only test except for the scores 12 and 26. The students in the upper ability group who obtained scores over 31 did better in the bilingual test. Generally, the LEP students' mathematics achievement is better in the bilingual test than the English-only test which suggests that the LEP students benefitted from the bilingual test as a test accommodation.

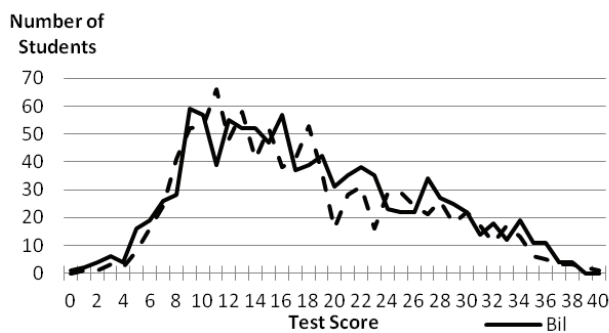


FIGURE 1. Score distribution for LEP students' total score in both tests

TABLE 2. Postsmoothing raw score moments for LEP students' total score

Test Form	Mean	Standard Deviation	Skewness	Kurtosis
English-only	17.5905	7.9558	0.5524	2.4127
Bilingual	17.9657	8.2127	0.4051	2.3203
Unsmoothed	17.5827	7.9499	0.5522	2.4108
S=0.01:	17.5834	7.9475*	0.5415*	2.3922
S=0.05:	17.5864	7.9460	0.5396	2.3883
S=0.10:	17.5869*	7.9451	0.5388	2.3893
S=0.20:	17.5949	7.9273	0.5340	2.4209
S=0.30:	17.6183	7.9178	0.5146	2.4212
S=0.40:	17.6425	7.9175	0.4938	2.4069*
S=0.50:	17.6628	7.9189	0.4769	2.3926
S=0.75:	17.7023	7.9236	0.4440	2.3631
S=1.00:	17.7334	7.9285	0.4182	2.3401
Linear:	17.5905	7.9558	0.4051	2.3203

Table 2 shows the results that were obtained when the scores of the two tests were linked by using equipercentile method. The value of $S=0.01$ represents the best approximation for the linked form because this value is the closest to the values of the mean, standard deviation, skewness and kurtosis for the English-only test. When using $S=0.01$, the linked scores for both tests are displayed in Table 3.

TABLE 3. Postsmoothing raw-to-raw score conversion for LEP students' total score

Bilingual Score	Unrounded English Score	Rounded English Score	Bilingual Score	Unrounded English Score	Rounded English Score
0	0.28	0	21	19.96	20
1	1.85	2	22	21.29	21
2	3.41	3	23	22.57	23
3	4.69	5	24	23.68	24
4	5.35	5	25	24.53	25
5	6.04	6	26	25.33	25
6	6.83	7	27	26.31	26
7	7.58	8	28	27.49	27
8	8.26	8	29	28.58	29
9	9.09	9	30	29.59	30
10	10.05	10	31	30.46	30
11	10.83	11	32	31.34	31
12	11.59	12	33	32.30	32
13	12.48	12	34	33.28	33
14	13.40	13	35	34.36	34
15	14.38	14	36	35.81	36
16	15.42	15	37	37.48	37
17	16.44	16	38	38.34	38
18	17.30	17	39	39.20	39
19	18.07	18	40	40.07	40
20	18.88	19			

Table 3 shows that the scores generally increased by one unit or increased by two units, decreased by one unit or there were no changes between the two linked scores. Table 4 was built to show the summarised changes between the two scores after linking was done.

TABLE 4. Summary of the raw-to-raw score conversion for LEP students' score

Raw Score Range for Bilingual Test	Change in LEP Students' Linked Score for English-only Test
8-12	No changes
13-22	-1
23-25	No changes
26-28	-1
29-30	No changes
31-35	-1

From Table 4, it can be seen that for the LEP students, the bilingual test was approximately one unit easier when compared to the English-only test. The lower end values and the upper end values are ignored as they tend to be less reliable due to smaller sample of students that causes large equating error as shown in Table 5.

As can be seen in Table 5, for the low scores the equating error ranges from 0.9753 to 1.8543 while for high scores the equating error ranges from 0.000 to 1.0684. For the score range of 8 to 35, the equating error range from 0.2697 to 1.2772. In view of the one unit score difference that is obtained between the scores of the two tests, the equating error that is obtained is rather large. As such, the one unit score difference may be due to equating error and as such, the bilingual test did not appear to offer assistance to the LEP students. The bilingual test had undergone processes to ensure good psychometric properties but not sufficient as a useful test accommodation that can alleviate the linguistics impediments of LEP students.

Even though the study focuses on exploring the validity of using the bilingual test in assessing LEP students' mathematics achievement, the scores of the two tests were also linked for the non-LEP students. This was done mainly to compare the mathematics achievement between the two groups of students in both tests. Table 6 shows the results when the non-LEP students' scores from the two tests were linked.

TABLE 5. Equating error for LEP students' total score

Score	sc	Score	sc	Score	sc	Score	sc
0	0.8022	11	0.2697	21	1.2772	31	0.5366
1	1.8543	12	0.4024	22	0.7095	32	0.6271
2	0.9753	13	0.3490	23	0.6201	33	0.8727
3	0.5009	14	0.3585	24	0.6355	34	0.4743
4	0.6305	15	0.5198	25	0.6201	35	0.4900
5	0.3931	16	0.4128	26	0.5937	36	0.9538
6	0.3419	17	0.5683	27	0.6804	37	1.0684
7	0.2491	18	0.5214	28	0.5862	38	0.8767
8	0.2905	19	0.3955	29	0.5437	39	0.0000
9	0.2705	20	0.5762	30	0.5825	40	0.0000
10	0.3094						

TABLE 6. Postsmoothing raw score moments for non-LEP students' total score

Test Form	Mean	Standard Deviation	Skewness	Kurtosis
English-only	26.7530	8.0628	-0.4713	2.3502
Bilingual	26.4170	8.4132	-0.4628	2.2891
Unsmoothed	26.7538	8.0582	-0.4703	2.3442
S=0.01:	26.7520	8.0630*	-0.4751	2.3572
S=0.05:	26.7524*	8.0646	-0.4757	2.3581
S=0.10:	26.7505	8.0709	-0.4768	2.3544*
S=0.20:	26.7396	8.0984	-0.4719*	2.3164
S=0.30:	26.7304	8.1037	-0.4655	2.3024
S=0.40:	26.7304	8.1037	-0.4655	2.3024
S=0.50:	26.7304	8.1037	-0.4655	2.3024
S=0.75:	26.7304	8.1037	-0.4655	2.3024
S=1.00:	26.7304	8.1037	-0.4655	2.3024
Linear:	26.7530	8.0628	-0.4628	2.2891

From Table 6, S=0.01 was found to be the most appropriate value due to the small difference between the central moments of the two tests. When using S=0.01, the linked scores for both

tests are displayed in Table 7 and it can be deduced that the linked scores either increased by one or there were no changes.

TABLE 7. Postsmoothing raw-to-raw score conversion for non-LEP students' total score

Bilingual Score	Unrounded English Score	Rounded English Score	Bilingual Score	Unrounded English Score	Rounded English Score
0	0.03	0	21	21.62	22
1	1.10	1	22	22.40	22
2	2.16	2	23	23.18	23
3	3.23	3	24	24.05	24
4	4.29	4	25	25.07	25
5	5.36	5	26	26.30	26
6	6.42	6	27	27.50	28
7	7.51	8	28	28.45	28
8	8.63	9	29	29.35	29
9	9.73	10	30	30.25	30
10	10.88	11	31	31.11	31
11	12.05	12	32	32.05	32
12	13.09	13	33	33.02	33
13	14.04	14	34	34.00	34
14	15.02	15	35	34.96	35
15	16.03	16	36	35.91	36
16	17.04	17	37	36.85	37
17	18.02	18	38	37.79	38
18	18.95	19	39	38.91	39
19	19.86	20	40	39.97	40
20	20.78	21			

The summary of the changes in the score is as exhibited in Table 8. Results indicated that the non-LEP students found the English-only test easier with a score difference of also one. The increase of one score at the lower range in the English-only test indicates that the English-only test is still easier for the non-LEP students. At the mid score range and higher end score range, there are no changes to the scores which indicate that the bilingual test did not help in their mathematics achievement.

TABLE 8. Summary of the raw-to-raw score conversion for non-LEP students' total score

Raw Score Range for Bilingual Test	non-LEP Students' Linked Score for English-only Test
7-21	+ 1
22-26	No changes
27	+ 1
28-40	No changes

For further analysis, the size of the equating error was inspected. Table 9 shows the values of the equating error. The equating error is exceptionally large for the score range of zero to seven as the equating errors are between the ranges of 0.7174 to 2.6675. For the lower range scores of 7 to 21, the equating error is relatively large when compared to an increase of one score unit after test linking. As such, it is highly possible that the one unit score difference is due to equating error and that the bilingual test did not assist the non-LEP students.

STUDENTS' RESPONSES ON THE UTILITY OF THE BILINGUAL TEST

Analysis revealed that both the LEP and non-LEP students relied on the Malay translation with 89.5% of the LEP students and 70.9 % of the non-LEP students found the Malay translation helpful while 88.6% of the LEP and 67.1% of the non-LEP students used it to understand the questions. Among the LEP students, 92% used the Malay translation to answer all the items. However in comparison, more LEP students depended on the items that were translated into the Malay language. The Malay translation helped the LEP students to understand the items in their attempt to answer them.

An interesting finding is that some of the LEP students claimed that they did not find the Malay translation useful because they were comfortable with the items presented in the English language as the language of instruction during their Mathematics lesson was also in English. Their simple comment like "teacher teaches in English, so I answer in English" speaks volume on the importance of the language of instruction to match the language of assessment despite their restricted English language proficiency. Since the language of assessment matched the language of instruction, there was no need to rely on the Malay language translation even though to some of them, it was their mother tongue.

TABLE 9. Equating error for the non-LEP students' total scores

Score	se	Score	se	Score	se	Score	se
0	2.0518	11	0.7049	21	0.4797	31	0.3466
1	2.0518	12	0.5552	22	0.4942	32	0.3295
2	2.0518	13	0.7608	23	0.4342	33	0.3511
3	0.8299	14	0.5107	24	0.4605	34	0.3335
4	2.4919	15	0.5765	25	0.5275	35	0.2775
5	2.6675	16	0.6389	26	0.5594	36	0.2961
6	0.8032	17	0.6153	27	0.4074	37	0.2250
7	0.7174	18	0.4501	28	0.4084	38	0.3014
8	0.5193	19	0.7028	29	0.4316	39	0.2504
9	0.3808	20	0.5143	30	0.4330	40	0.1954
10	0.5194						

Another noteworthy finding is that among the LEP students who were from the Chinese schools, they did not find the bilingual test especially the Malay translation helpful. This is because they had received their primary education in the Chinese language and later on in the English language in the secondary school and as such, the Malay translation was of no help as it was not the language of instruction at both levels of their education.

DISCUSSION AND CONCLUSION

The one score increase in the bilingual test for LEP students may be due to equating error and therefore, the bilingual test was not easier for the LEP students. However when analysing the students' responses on the utility of the Malay version, it can be concluded that the bilingual test helped to arrest their language deficiency. The Malay language version was found to have assisted certain groups of LEP students to understand the questions. However, they still preferred to use the mathematical terminologies in the language of instruction which was English. These LEP students probably used the Malay translation to understand the linguistics and syntax components of the questions but not the mathematical terminologies in English. This is because these terminologies that they were exposed to them during their mathematics classes were different from the mathematical terminologies in the Malay language. Therefore, despite facing difficulties in understanding the items in English, they still did not depend on the adapted items in Malay. As such, it is possible to claim that simplifying the linguistics of the items in English may benefit the LEP students even better and future studies can be directed towards this test accommodation before further claims can be made.

In addition, the LEP students who had received instruction at the primary level in Chinese language did not benefit as much as the LEP students whose language of instruction at the primary level was in the Malay language. Even though the Malay language was also their dominant language as it is the national language, the importance of the Chinese language as the language of instruction during their Mathematics lesson at the primary level was more overpowering. This is because students were more familiar with the mathematical terminologies used

in that language during instruction. These results seem to suggest that the bilingual test would have yielded more promising results if they had been tested in two languages that had been the instructional languages for all groups of students at both the primary and secondary levels. Bilingual test helps only if the adapted items are written in the language of instruction and another proficient language which was also used during instruction.

In addition, the findings concur that the bilingual test did not unnecessarily help the non-LEP students. As Sireci (1997) reiterated, test accommodation should not put one group at an either advantage or disadvantage. These students used the bilingual items in Malay to understand the items written in English as was found by Duncan et al. (2005) where the bilingual test items were used as to arrest the language complications inherent in the items.

In the Malaysian context, the language of assessment should match the language of instruction despite students' limited English language proficiency. This study, therefore concurs to the findings of previous studies (Abedi 2006; Chamberlain 2005; Hofstter 2003). Just like the findings of studies done elsewhere (Abedi & Lord 2001; Abedi 2002), the findings of this study also point along the same direction that linguistics simplification of test items is promising. By simplifying the linguistics features of the items in English, it will greatly benefit the LEP students (Abedi & Hejri 2004; Abedi 2006) because the linguistics complexity that contributes to the construct-irrelevant variance is removed (Abedi 2006).

Linguistic simplification fulfils two important aspects of testing which are it allows the language of assessment to match the language of instruction which does not alter the psychometric properties of a test and secondly, it removes the language obstacle that burdens LEP students.

IMPLICATIONS AND LIMITATIONS OF THE RESEARCH FINDINGS

The implications of this research finding bear importance in the Malaysian national assessments. Being proficient in the language of test may ease in understanding the linguistics features of the questions, but in order to solve the Mathematics

questions, an in-depth understanding of the mathematical terminologies in the language of instruction is not only necessary but a prerequisite. Therefore, the language of assessment must also be the language of instruction so that students are very familiar due to the high exposure during the Mathematics lesson. In national assessments, linguistic simplification is a noteworthy consideration that must be practised especially during test item construction as the unnecessary language load is reduced. Test item writers should use simple language when developing Mathematics test items as well as the test instructions. This is to reduce, if not remove language as a construct irrelevant variance that violates test validity due to emergence of language as a secondary dimension that is being assessed as a part of the test construct.

The findings also shed light in the usefulness of the bilingual test among LEP students and the examines the challenges (if any) faced by them when answering in a language that is not their native language.

The main limitation of this finding is that it should only be interpreted within the context of this study where the language of instruction was in English and the items that were used were also in the English language. This finding should not be generalised to future policy which reverts the language of instruction for Mathematics to the Malay language. This is because the Malay language is the national language and is understood by all Malaysian students as it is formally taught in all Malaysian schools.

REFERENCES

- Abedi, J. 2006. Psychometric issues in the ELL assessment and special education eligibility. *Teachers College Record* 108(11): 2282-2303.
- Abedi, J., Courtney, M., Leon, S., Kao, J & Azzam, T. 2006. *English Language Learners and Math Achievement: A Study of Opportunity to Learn and Language Accommodation*. Technical Report 702. CRESST.
- Abedi, J. 2004. *Inclusions of Students with Limited English Proficiency in NAEP: Classifications and Measurement Issues*. Retrieved June 26, 2007, from <http://www.nagb.org/pubs/conferences/abedi.pdf>.
- Abedi, J. & Hejri, F. 2004. Accommodations for students with limited English proficiency in the National Assessment of Educational Progress. *Journal of Applied Measurement in Education* 17(4): 371 - 392.
- Abedi, J. 2002. Assessments and accommodation of English language: Issues, concerns and recommendations. *Journal of School Improvement* 3(1): 83-89.
- Abedi, J. & Lord, C. 2001. The language factor in Mathematics test. *Applied Measurement in Education* 14(3): 219-234.
- Ain Nadzimah Abdullah & Chan, S.W. 2003. Gaining linguistic capital through a bilingual language policy innovation. *South Asian Language Review* XIII:1-2.
- Ainan Abdul Samad. 2003. *English as a Tool for Science and Technology Knowledge Acquisition: Bilingual Assessment Instruments as a Transitional Measure*. Paper presented at the International Association of Educational Assessment, October 5-10, Manchester, United Kingdom.
- Althouse, L.A. 2000. *Test Development: Ten Steps to a Valid and Reliable Certification exam*. Paper presented at the Proceedings of the Twenty-fifth Annual: SAS Users Group International Conference, April 9-12, Indianapolis, Indiana.
- American Educational Research Association, American Psychological Association & National Council on Measurement in Education. 1999. *Standard for Educational and Psychological Testing*. Washington, DC: American Psychological Association
- Bernardo, A.B.I. & Calleja, M.O. 2005. The effects of stating problems in bilingual students' first and second languages on solving mathematical word problem. *The Journal of Genetic Psychology* 166(1): 117-128.
- Center for Excellence and Equity in Education. 2005. *Assessment and Accountability: Accommodation*. Retrieved August 29, 2007, from <http://ceee.gwu.edu/AA/Accommodations.html>.
- Duncan, T.G. et al. 2005. Study of a dual-language test booklet in eighth-grade Mathematics. *Journal of Applied Measurement in Education* 18(2): 129-161.
- Educational Testing Service. 2002. *ETS Standards for Quality and Fairness*. Retrieved June 2, 2007, from http://www.ets.org/Media/About_ETS/pdf/standards.pdf.
- Francis, D.J., Rivera, M., Lesaux, N., Kieffer, M. & Rivera, H. (2006). *Practical guidelines for the education of English language learners: Research-based recommendations for the use of accommodations in large-scale assessments*. Retrieved June 23, 2007, from www.centreoninstruction.org/files/ELL3_Assessments.pdf
- Grisay, A. 2003. Translation procedures in OECD/PISA 2000 international assessment. *Language Testing* 20(2): 225-240.
- Hofstetter, C.H. 2003. Contextual and Mathematics accommodation test effects for English language learners. *Applied Measurement in Education* 16(2): 159-188. Retrieved October 1, 2008, from <http://www.informaworld.com/smpp/content~content=a783684874~db=all>.
- Holland, P.W. & Dorans, N.J. 2006. Linking and equating. In *Educational measurement*, (edited by Arennan. R.L. United States of America: Praeger Publishers.

- Hulin, C.L. 1987. A psychometric theory of evaluations of item and scale translation: Fidelity Across Language. *Journal of Cross-Cultural Psychology* 18(2): 115-142.
- Kolen, M.J. 2005. *RAGE-RGEQUATE (Version 3.22)* [Computer Software]. The University of Iowa, Iowa City: Iowa Testing Programs.
- Kolen, M.J. & Brennan, R.L. 1995. *Test Equating: Methods and Practices*. Iowa City, USA: Springer-Verlag New York, Inc.
- Lim, C.S. 2002. *Cultural Differences and Mathematics Learning in Malaysia*. [Electronic Version]. Retrieved April 24, 2007, from <http://www.math.admu.edu/tsg22/lim.html>.
- Linacre, J.M. 1994. *What Do Infit and Outfit, Mean Square and Standardized Mean*. Retrieved May 29, 2008, from <http://www.Rasch.org/rmt/rmt162f.htm>.
- National Association for the Education of Young Children. 2005. *Screening and Assessments of Young English-language Learners*. Retrieved May 25, 2007, from http://www.naeyc.org/about/positions/ELL_Supplement.asp.
- Pena, E.D. 2007. Lost in translation. Methodological considerations in cross-cultural research. *Journal of Child Development* 78 (4): 1255-1264.
- Rapp, J., & Allalouf, A. (2003). Evaluating cross-lingual equating. *International Journal of Testing*, 3(2): 101-117.
- Schmeiser, C.B. & Welsch, C.J. 2006. Test development. In *Educational measurement*, edited by Brennan, R.L. United States of America: Praeger Publishers.
- Sireci, S.G. 1997. Problems and issues in linking tests across languages. *Journal of Educational Measurement* 16(1): 12-19.
- Stansfield, C.W. 1997. *Experience and Issues Related to the Format of Bilingual Tests: Dual Language Test Booklets Versus Two Different Test Booklets*. Retrieved May 12, 2007 from <http://www.eric.ed.gov/PDFS/ED419002.pdf>.
- Thurlow, M.L., McGrew, K. S., Tindal, G. , Thompson, S.L., Yoseldyke, J. E. & Elliot, J.L. 2000. *Assessment Accommodations Research: Considerations for Design and Analysis*. NCEO Technical Report 26. National Centre on Educational Outcomes. Retrieved September, 11, 2007 from <http://education.umn.edu/NCEO/OnlinePubs/Technical26.htm>.
- Messick, S.C. 1995. A validity of psychological assessment: Validation from persons' responses and performances as scientific inquiry information score meaning. *American Psychologist* 50(9), 741-749.
- Ministry of Education. 2004. *The Development of Education, National Report of Malaysia*. Retrieved December 15, 2007, from <http://www.ibe.unesco.org/International/ICE47/english/Natreps/reports/malaysia.pdf>.
- Mullis, I.V.S., Martin, M.O. & Diaconu, D. (2004). Item analysis and review. In Martin, M.O., Mullis, I.V.S., & Chrostowski, S. J, *TIMSS 2003 Technical Report*, editor by Chesnut Hill, MA: TIMSS and PIRLS International Study Centre, Boston College. Retrieved November 2, 2007, from <http://timss.bc.edu/timss2003i/technicalD.html>
- Wilde, J. 2007. *Resources About Accommodations for English Language Learners*. National Clearinghouse for English language acquisition and language instruction educational programs. Retrieved June 1, 2007, from <http://www.nclea.gwu.edu/resabout/accomodations/intro/index.html>.
- Wright, B. & Linacre. 1994. *A Rasch Unidimensionality coefficient*. [Electronic Version]. Retrieved January 2, 2008, from <http://www.rasch.org/rmt/rmt83p.htm>.
- S.Kanageswari d/o Suppiah Shanmugam*
Southeast Asian Ministers of Education Organisation (SEAMEO)
Regional Centre for Education in Science and Mathematics (RECSAM),
Penang, Malaysia
- Ong Saw Lan
Universiti Sains Malaysia
Penang
- *Corresponding author;
email: kanageswari@recsam.edu.my
- Received : 3rd January 2012
Accepted : 1st February 2013